

PEMETAAN KLASTER PEMBANGUNAN KABUPATEN/KOTA DI PROVINSI BALI BERBASIS *DEEP CLUSTERING* MENGGUNAKAN *VARIATIONAL AUTOENCODER* DAN *IMPROVED K-MEANS*

Kanessa Jasmine Prisheila Az Zahra Setiyanto, Amri Muhaimin, Kartika Maulida Hindrayani

Sains Data, Universitas Pembangunan Nasional “Veteran” Jawa Timur

Jl. Raya Rungkut Madya, Gunung Anyar, Surabaya

amri.muhamin.stat@upnjatim.ac.id

ABSTRAK

Ketimpangan pembangunan antar kabupaten/kota menimbulkan perbedaan capaian sosial ekonomi sehingga diperlukan pemetaan pembangunan wilayah secara objektif. Penelitian ini bertujuan mengelompokkan pembangunan kabupaten/kota di Provinsi Bali periode 2020-2024 berdasarkan indikator Indeks Pembangunan Manusia (IPM), Produk Domestik Regional Bruto (PDRB), Gini Ratio (GR), Tingkat Pengangguran Terbuka (TPT), dan persentase penduduk miskin (P0). Metode yang digunakan merupakan metode *deep clustering* dengan *Variational Autoencoder* (VAE) untuk membentuk representasi fitur laten, yang selanjutnya dikeompokkan menggunakan *Improved K-Means Clustering*. Penentuan jumlah kluster dilakukan dengan menguji $K = 3, 4, 5$ yang dievaluasi menggunakan *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CHI). Hasil penelitian menunjukkan bahwa $K = 3$ merupakan jumlah kluster yang paling optimal, dengan kualitas kluster terbaik terjadi pada tahun 2020 dan 2023, ditunjukkan dengan nilai *Silhouette* masing-masing sebesar 0,467 dan 0,390. Profil indikator indikator menunjukkan bahwa kluster dengan tingkat pembangunan lebih tinggi memiliki nilai IPM dan PDRB per kapita yang lebih tinggi serta tingkat kemiskinan lebih rendah, yang didominasi oleh Kabupaten Badung dan Kota Denpasar. Hasil ini diharapkan dapat mendukung analisis dan perumusan kebijakan pembangunan daerah di Provinsi Bali.

Kata kunci : *Kluster Pembangunan, Variational Autoencoder, Improved K-Means Clustering, Kabupaten/Kota, Provinsi Bali.*

1. PENDAHULUAN

Pembangunan wilayah merupakan proses yang bertujuan meningkatkan kesejahteraan masyarakat melalui pemerataan akses layanan dasar, peningkatan kualitas hidup, serta pertumbuhan ekonomi [1]. Namun, dalam praktiknya pembangunan antar wilayah tidak selalu berjalan merata sehingga memunculkan ketimpangan sosial ekonomi. Ketimpangan tersebut dapat diamati melalui perbedaan capaian sejumlah indikator pembangunan seperti Indeks Pembangunan Manusia (IPM), Produk Domestik Regional Bruto (PDRB), Tingkat Pengangguran Terbuka (TPT), persentase penduduk miskin (P0), dan Gini Ratio (GR) [2]. Perbedaan capaian indikator ini menunjukkan bahwa setiap wilayah memiliki karakteristik pembangunan yang berbeda sehingga diperlukan pemetaan pembangunan wilayah sebagai dasar evaluasi dan perumusan kebijakan yang lebih tepat sasaran.

Provinsi Bali dipilih sebagai objek penelitian karena memiliki karakteristik pembangunan yang unik, yaitu adanya konsentrasi aktivitas ekonomi dan pusat pertumbuhan pada wilayah tertentu yang berpotensi menimbulkan ketimpangan antar kabupaten/kota. Selain itu, periode 2020–2024 merupakan fase yang penting untuk dianalisis karena mencerminkan dinamika sosial ekonomi daerah dalam beberapa tahun terakhir. Kondisi tersebut menimbulkan permasalahan utama berupa bagaimana memetakan dan mengelompokkan kabupaten/kota di Provinsi Bali secara objektif berdasarkan kemiripan

indikator pembangunan agar perbedaan karakteristik pembangunan antar wilayah dapat diidentifikasi dengan jelas.

Salah satu pendekatan yang dapat digunakan untuk menyelesaikan permasalahan tersebut adalah metode *clustering*, yang bertujuan mengelompokkan wilayah berdasarkan kemiripan karakteristik indikator. Namun, metode *clustering* konvensional seperti K-Means memiliki keterbatasan, antara lain sensitif terhadap perbedaan skala data dan pemilihan centroid awal yang dapat memengaruhi stabilitas hasil klusterisasi [3]. Oleh karena itu, diperlukan pendekatan yang lebih adaptif untuk menghasilkan kluster yang lebih representatif.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan pendekatan *deep clustering* dengan memanfaatkan *Variational Autoencoder* (VAE) dan *Improved K-Means*. VAE digunakan untuk membentuk representasi data pada ruang laten (*latent space*) yang lebih informatif, sedangkan *Improved K-Means* digunakan untuk menghasilkan kluster yang lebih stabil melalui pembobotan fitur dan perhitungan jarak yang lebih adaptif. Penelitian ini bertujuan untuk mengelompokkan pembangunan kabupaten/kota di Provinsi Bali periode 2020–2024 berdasarkan indikator IPM, PDRB, GR, TPT, dan P0, serta mengevaluasi kualitas kluster menggunakan *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CHI).

Hasil pengelompokan divisualisasikan dalam bentuk peta kluster pembangunan untuk

menggambarkan persebaran dan dinamika pola pembangunan wilayah di Provinsi Bali. Kontribusi penelitian ini adalah penerapan pendekatan *deep clustering* berbasis VAE dan Improved K-Means untuk pemetaan kluster pembangunan daerah, serta interpretasi hasil kluster berdasarkan profil rata-rata indikator guna mengidentifikasi perbedaan karakteristik pembangunan relatif antar wilayah.

2. TINJAUAN PUSTAKA

Tinjauan pustaka pada penelitian ini membahas konsep dan teori yang mendasari proses pengelompokan pembangunan daerah. Pembahasan mencakup indikator pembangunan yang digunakan, konsep *Variational Autoencoder* (VAE) sebagai metode pembentukan representasi laten, *Improved K-Means* sebagai metode klusterisasi, metrik evaluasi kualitas kluster, serta interpretasi rata-rata tiap kluster untuk membedakan karakteristik pembangunan relatif antar kluster.

2.1. Indikator dan Karakteristik Pembangunan Daerah

Karakteristik pembangunan daerah menggambarkan kondisi sosial ekonomi suatu wilayah berdasarkan capaian indikator pembangunan yang digunakan. Dalam penelitian ini, karakteristik pembangunan kabupaten/kota direpresentasikan melalui lima indikator utama, yaitu Indeks Pembangunan Manusia (IPM), Produk Domestik Regional Bruto (PDRB), Gini Ratio (GR), Tingkat Pengangguran Terbuka (TPT), dan persentase penduduk miskin (P0). Pemilihan indikator tersebut dilakukan karena mampu merepresentasikan dimensi pembangunan manusia, kondisi ekonomi, ketimpangan, ketenagakerjaan, serta kemiskinan secara komprehensif.

Indeks Pembangunan Manusia (IPM) digunakan untuk menggambarkan kualitas pembangunan manusia suatu wilayah yang mencakup aspek kesehatan, pendidikan, dan standar hidup layak [4]. Nilai IPM yang lebih tinggi menunjukkan capaian pembangunan manusia yang lebih baik. Produk Domestik Regional Bruto (PDRB) mewakili indikator ekonomi yang menggambarkan tingkat kegiatan dan keluaran ekonomi daerah. PDRB yang lebih tinggi menunjukkan kemampuan ekonomi wilayah yang lebih kuat [5].

Gini Ratio (GR) merupakan indikator statistik yang dapat digunakan untuk mengukur tingkat ketimpangan distribusi pendapatan dalam suatu wilayah. Nilai GR yang semakin besar menunjukkan ketimpangan yang semakin tinggi [6]. Tingkat Pengangguran Terbuka (TPT) menggambarkan kondisi ketenagakerjaan melalui proporsi angkatan kerja yang belum terserap dalam pasar kerja. Nilai TPT yang tinggi mengindikasikan adanya tantangan dalam penyerapan tenaga kerja. Sementara itu, persentase penduduk miskin (P0) menunjukkan proporsi penduduk yang berada di bawah garis

kemiskinan, sehingga nilai P0 yang lebih tinggi mencerminkan tingkat kemiskinan yang lebih besar.

Kelima indikator tersebut digunakan sebagai dasar untuk membentuk pengelompokan wilayah dan menginterpretasikan perbedaan karakteristik pembangunan relatif. Secara umum, wilayah dengan karakteristik pembangunan relatif lebih baik cenderung memiliki nilai IPM dan PDRB yang lebih tinggi, serta nilai P0 dan Gini Ratio (GR) yang lebih rendah, sedangkan wilayah dengan karakteristik pembangunan relatif lebih rendah menunjukkan pola yang berlawanan. Dengan demikian, kombinasi indikator ini memberikan gambaran yang lebih komprehensif dalam menganalisis pola pembangunan daerah.

2.2. Variational Autoencoder (VAE)

Variational Autoencoder (VAE) merupakan model *deep learning* yang digunakan untuk mempelajari representasi data pada ruang laten (*latent space*) secara probabilistik [7]. VAE terdiri dari dua komponen utama, yaitu *encoder* dan *decoder*, yang bekerja untuk membentuk representasi laten sekaligus merekonstruksi kembali data input [8]. *Encoder* memetakan data input x ke dalam distribusi laten menggunakan persamaan (1).

$$q\phi(z|x) = N(z; \mu(x), \sigma^2(x)) \quad (1)$$

Kemudian sampel vektor variabel laten diambil dari distribusi tersebut dengan menggunakan:

$$z = \mu(x) + \sigma(x) \cdot \epsilon, \quad \epsilon \sim N(0,1) \quad (2)$$

Setelah itu, *decoder* memetakan variabel laten z kembali ke ruang data sehingga menghasilkan rekonstruksi $\hat{x}$. Proses rekonstruksi ini memastikan informasi penting dari data masukan tetap dapat dipertahankan melalui representasi laten yang dipelajari. Fungsi objektif VAE terdiri dari dua komponen utama, yaitu loss rekonstruksi dan regularisasi distribusi laten menggunakan *Kullback-Leibler divergence* (KL). Secara umum, fungsi loss VAE dapat dituliskan sebagai:

$$L = E_{q\phi(z|x)} [\log p\phi(x|z)] - D_{KL}(q\phi(z|x)||p(z)) \quad (3)$$

Komponen rekonstruksi mendorong model menghasilkan $\hat{x}$ yang mendekati data asli x , sedangkan komponen KL memastikan distribusi laten $q\phi(z|x)$ tetap terstruktur dan mendekati prior $p(z) = N(0, I)$.

2.3. Improved K-Means Clustering

Improved K-Means Clustering merupakan pengembangan dari K-Means yang memperbaiki penentuan centroid awal untuk mendapatkan hasil pengelompokan yang lebih konsisten. Melalui metode ini, proses clustering lebih adaptif terhadap karakteristik data serta dapat mengurangi kemungkinan terbentuknya kluster yang kurang optimal karena pemilihan centroid awal yang tidak stabil karena ditentukan secara acak [9]. Tahap

pertama dalam *Improved K-Means Clustering* adalah perhitungan bobot tiap fitur menggunakan variasi data dimana ditur ke- j dari nilai rata-rata dan standar deviasi dihitung dengan menggunakan persamaan (4) dan (5).

$$\mu_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (4)$$

$$\sigma = \frac{1}{\mu_j} \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \mu_j)^2} \quad (5)$$

Selanjutnya digunakan *Coefficient of Variation* (CV) untuk menggambarkan variasi relative dari data dengan rumus:

$$CV_j = \left| \frac{\sigma_j}{\mu_j} \right| \quad (6)$$

Nilai CV kemudian dinormalisasi sehingga diperoleh bobot fitur w_j dengan jumlah bobot sama dengan satu menggunakan persamaan (7).

$$w_j = \frac{CV_j}{\sum_{j=1}^p CV_j} \quad (7)$$

Setelah bobot diperoleh, kemiripan antar data dihitung menggunakan *Weighted Euclidean Distance* persamaan (8).

$$\text{dis}(x_a, x_b) = \sqrt{\sum_{j=1}^m W_j (x_{aj} - x_{bj})^2} \quad (8)$$

Jarak berbobot ini memungkinkan fitur yang memiliki kontribusi lebih besar memberikan pengaruh yang lebih dominan dalam pembentukan kluster. Kemudian inisialisasi centroid awal ditentukan dengan menggunakan konsep *nearest neighbors*. Dengan perhitungan jarak rata-rata antar data sebagai ambang batas menggunakan perhitungan (9).

$$d_{avg} = \frac{1}{n^2} \sum_{a=1}^n \sum_{b=1}^n d(x_a, x_b) \quad (9)$$

Kemudian, setiap data dihitung jumlah tetangga dekatnya, yaitu jumlah data lain yang memiliki jarak lebih kecil dari d_{avg} . Data dengan jumlah tetangga dekat terbanyak dipilih sebagai centroid pertama. Centroid selanjutnya dipilih dengan mempertimbangkan kandidat dengan jumlah tetangga dekat tinggi, namun kandidat yang berada terlalu dekat (menjadi tetangga dekat) dengan centroid yang sudah terpilih akan dilewati. Proses tersebut dilakukan hingga jumlah centroid awal sesuai dengan jumlah kluster yang ditentukan, sehingga centroid awal lebih tersebar dan hasil klusterisasi menjadi lebih stabil.

2.4. Evaluasi Kualitas Kluster

Tahap evaluasi dilakukan untuk menilai kluster yang terbentuk dengan memastikan bahwa anggota dalam satu kluster memiliki karakteristik yang sama,

sedangkan antar kluster memiliki perbedaan yang jelas. Pada penelitian ini digunakan tiga metrik evaluasi, yaitu *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CHI) untuk menentukan jumlah kluster terbaik serta menilai tingkat kekompakan dan pemisahan antar kluster sebelum dilakukan interpretasi hasil kluster [10].

2.5. Silhouette Score

Silhouette Score digunakan untuk mengevaluasi seberapa baik suatu objek ditempatkan dalam kluster tertentu, dengan membandingkan kedekatannya terhadap anggota dalam kluster yang sama dengan kedekatannya terhadap kluster lain yang paling berdekatan [11]. Semakin besar nilai yang didapatkan dari *Silhouette Score* (mendekati 1) mengindikasikan bahwa pengelompokan data kluster semakin baik [12]. Dalam penelitian ini, *Silhouette Score* dimanfaatkan untuk mengevaluasi hasil klusterisasi pada setiap tahun sekaligus mendukung pemilihan jumlah kluster yang paling sesuai. *Silhouette Score* dirumuskan pada persamaan (10).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (10)$$

2.6. Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) digunakan untuk mengukur performa kluster berdasarkan hubungan antara tingkat penyebaran data di dalam kluster dan jarak pemisah antar kluster [13]. Secara umum, DBI akan menghasilkan nilai yang lebih kecil apabila kluster semakin kompak dan jarak antar centroid semakin jauh. Oleh karena itu, semakin rendah nilai DBI maka semakin baik kualitas pengelompokan yang dihasilkan. Pada penelitian ini, DBI dihitung pada setiap hasil klusterisasi tahunan untuk melihat kestabilan kualitas kluster dari waktu ke waktu. Rumus DBI dituliskan sebagai:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (11)$$

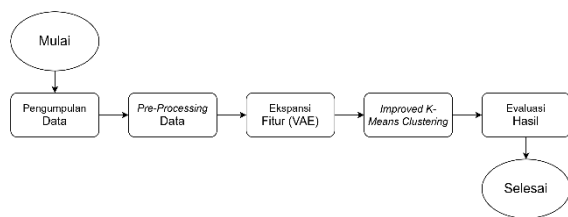
2.7. Calinski-Harabasz Index (CHI)

Calinski-Harabasz Index (CHI) digunakan sebagai ukuran tambahan untuk menilai kualitas kluster berdasarkan perbandingan variasi antar kluster terhadap variasi di dalam kluster [14]. Nilai CHI yang tinggi mengindikasikan pemisahan antar kluster semakin jelas, sementara penyebaran data di dalam kluster relatif lebih kecil. Dalam penelitian ini, CHI digunakan untuk memperkuat hasil evaluasi *Silhouette Score* dan DBI sehingga pemilihan konfigurasi kluster dapat dilakukan secara lebih meyakinkan. Rumus CHI dinyatakan sebagai berikut:

$$CHI = \frac{\text{Tr}(B_k)}{\text{Tr}(W_k)} \times \frac{n - K}{K - 1} \quad (12)$$

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan deep clustering untuk mengelompokkan kabupaten/kota di Provinsi Bali berdasarkan indikator pembangunan periode 2020–2024. Tahapan penelitian meliputi pengumpulan data, *pre-processing* data, ekspansi fitur menggunakan *Variational Autoencoder* (VAE) dan dilanjutkan dengan melakukan clustering menggunakan *Improved K-Means Clustering*. Metrik evaluasi seperti *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CHI) digunakan untuk menilai kualitas hasil clustering. Alur penelitian secara umum ditunjukkan pada Gambar 1.



Gambar 1. Diagram alir penelitian

3.1. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder periode 2020–2024 dari 9 wilayah kabupaten/kota yang berada di Provinsi Bali yaitu Jembrana, Tabanan, Badung, Gianyar, Klungkung, Bangli, Karangasem, Buleleng, dan Kota Denpasar. Indikator pembangunan yang digunakan terdiri dari lima variabel utama, yaitu Indeks Pembangunan Manusia (IPM), Produk Domestik Regional Bruto (PDRB), Gini Rasio (GR), Tingkat Kemiskinan (P0), dan Tingkat Pengangguran Terbuka (TPT). Selanjutnya, data disusun dalam bentuk tabel per tahun, dimana setiap baris merepresentasikan satu kabupaten/kota dan setiap kolom merepresentasikan indikator pembangunan. Data yang telah tersusun kemudian digunakan sebagai input pada tahap *pre-processing* dan pemodelan.

3.2. Pre-processing Data

Tahap *pre-processing* data dilakukan guna memastikan data dapat diproses dengan baik pada tahapan selanjutnya. Hal ini dilakukan karena setiap indikator memiliki skala dan satuan yang berbeda. Oleh karena itu, data dinormalisasi agar salah satu variabel tidak mendominasi. Pada penelitian ini, normalisasi dilakukan secara terpisah untuk setiap tahun menggunakan *RobustScaler*. Metode ini dipilih karena mampu mengurangi pengaruh outlier dengan memanfaatkan nilai median dan *interquartile range* (IQR) [15]. Pendekatan ini bertujuan menjaga konsistensi analisis pada masing-masing tahun serta menghindari bias akibat perbedaan skala indikator antar waktu. Data hasil normalisasi kemudian digunakan sebagai input untuk proses ekstraksi fitur menggunakan *Variational Autoencoder* (VAE).

3.3. Ekspansi Fitur Menggunakan *Variational Autoencoder* (VAE)

Setelah dilakukan proses *pre-processing* yang berupa normalisasi data, dilakukan tahapan selanjutnya yakni ekspansi fitur menggunakan *Variational Autoencoder* (VAE). VAE digunakan untuk mempelajari representasi data pada ruang laten (*latent space*) yang dapat menangkap hubungan non-linear antar indikator pembangunan untuk membentuk fitur baru yang lebih kaya. Dalam penelitian ini, 5 variabel yang ada dalam data awal penelitian diekspansi menjadi 10 fitur hasil pembelajaran VAE. Output dari tahap ini adalah *encoder* berupa vektor laten Z yang kemudian digunakan sebagai input utama pada tahap clustering menggunakan *Improved K-Means*.

3.4. Clustering Menggunakan *Improved K-Means Clustering*

Setelah mendapatkan vektor laten Z dari hasil ekspansi fitur menggunakan VAE, dilakukan pengelompokan kabupaten/kota di Provinsi Bali menggunakan metode *Improved K-Means Clustering*. Metode *Improved K-Means Clustering* merupakan pengembangan dari algoritma *K-Means* standar yang bertujuan untuk menghasilkan kluster yang lebih stabil dan representative. Proses clustering dilakukan untuk setiap tahun pada tahun 2020–2024 menggunakan fitur laten Z sebagai input. Hasil dari tahap ini adalah pengelompokan kabupaten/kota ke dalam beberapa kluster yang selanjutnya diinterpretasikan untuk mengidentifikasi perbedaan karakteristik pembangunan relatif antar wilayah.

3.5. Evaluasi Hasil Clustering

Evaluasi hasil clustering dilakukan untuk menilai kualitas hasil pengelompokan kabupaten/kota di Provinsi Bali pada periode 2020–2024. Evaluasi dilakukan untuk setiap tahun secara terpisah, sehingga kualitas kluster dapat dibandingkan antar periode. Metrik evaluasi yang digunakan dalam proses evaluasi adalah *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CHI). *Silhouette Score* digunakan untuk mengukur tingkat kekompakan dan pemisahan antar kluster, DBI untuk melihat ketimpang tindihan antar kluster, sedangkan CHI untuk mengukur rasio pemisahan antar kluster terhadap kepadatan yang ada dalam kluster.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil, pengujian, dan pembahasan yang meliputi statistika deskriptif data, penentuan jumlah kluster, evaluasi kualitas kluster, serta pemetaan hasil kluster untuk menggambarkan pola dan karakteristik pembangunan wilayah.

4.1. Statistika Deskriptif Data

Statistika deskriptif digunakan untuk memberikan gambaran umum mengenai karakteristik data indikator pembangunan kabupaten/kota di Provinsi Bali pada periode 2020–2024. Pada tahap awal ini dilakukan analisis awal berupa nilai rata-rata, standar deviasi, nilai minimum, dan nilai maksimum dari setiap variabel yang digunakan, yaitu PDRB, IPM, GR, TPT, dan P0. Hasil perhitungan statistika deskriptif ditunjukkan pada Tabel 1.

Tabel 1. Statistika Deskriptif

Variabel	Mean	Standar deviasi	Min	Max
PDRB	51605.21	23168.31	25999.00	132087.94
IPM	74.51	4.45	67.35	83.87
GR	0.31	0.02	0.28	0.36
TPT	3.48	2.00	0.74	7.53
P0	4.90	1.22	2.02	6.98

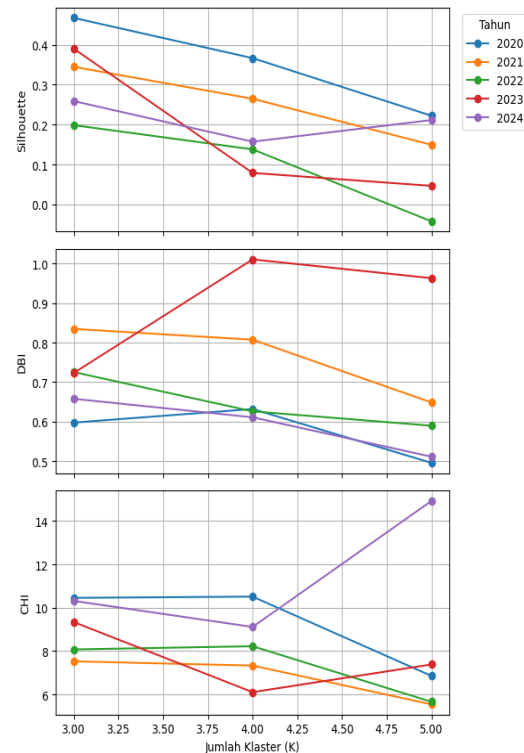
Dari Tabel 1 dapat dilihat bahwa indikator PDRB memiliki nilai rata-rata sebesar 51.605,21 dengan standar deviasi 23.168,31 perbedaan besar dapat dilihat pada nilai minimum dan maksimum yang berada pada rentang nilai 25.999,00 – 132.087,94 menunjukkan perbedaan aktivitas ekonomi yang signifikan antar kabupaten/kota di Provinsi Bali. Indikator IPM memiliki rata-rata 74,51 dalam rentang 67,35 sebagai nilai minimum dan 83,87 pada nilai maksimum, mengindikasikan adanya variasi capaian pembangunan manusia antar wilayah. Sementara itu, nilai Gini Ratio (GR) berada pada rata-rata 0,31 dengan nilai minimum 0,28 dan 0,36 nilai maksimum, menunjukkan tingkat ketimpangan pendapatan pada kategori moderat.

Pada indikator sosial ekonomi lainnya, TPT memiliki rata-rata 3,48 dengan rentang 0,74–7,53, sedangkan P0 memiliki rata-rata 4,90 dengan rentang 2,02 – 6,98. Variasi nilai pada kedua indikator tersebut menunjukkan adanya perbedaan kondisi ketenagakerjaan dan tingkat kemiskinan antar kabupaten/kota di Provinsi Bali selama periode pengamatan. Rentang nilai yang cukup lebar pada kedua indikator tersebut mengindikasikan bahwa dinamika sosial ekonomi tiap wilayah tidak seragam dan dapat membentuk karakteristik pembangunan yang berbeda antar daerah.

4.2. Penentuan Jumlah Kluster Optimal

Penentuan jumlah kluster dilakukan dengan menguji nilai $K = 3, 4$, dan 5 menggunakan metode *Improved K-Means* pada fitur latent Z di setiap tahun 2020–2024. Pemilihan K dievaluasi menggunakan *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CHI). Hasil evaluasi menunjukkan bahwa nilai *Silhouette* cenderung paling tinggi pada $K=3$, sedangkan nilai DBI dan CHI pada beberapa tahun menunjukkan perubahan yang bervariasi ketika nilai K ditingkatkan. Visualisasi perbandingan ketiga metrik evaluasi untuk setiap nilai K ditunjukkan pada Gambar 2.

Evaluasi Pemilihan Jumlah Kluster (K) Berdasarkan Silhouette, DBI, dan CHI (2020–2024)



Gambar 2. Perbandingan K terbaik

Berdasarkan grafik evaluasi pemilihan jumlah kluster tahun 2020–2024, nilai *Silhouette Score* secara konsisten paling tinggi pada $K=3$, yaitu sebesar 0,467 (2020), 0,345 (2021), 0,199 (2022), 0,390 (2023), dan 0,259 (2024). Hal ini menunjukkan bahwa pembentukan kluster pada $K=3$ menghasilkan struktur kluster yang lebih kompak dan pemisahan yang lebih jelas dibandingkan $K=4$ dan $K=5$. Pada metrik *Davies-Bouldin Index* (DBI), penurunan nilai pada K yang lebih besar hanya terjadi pada beberapa tahun, seperti tahun 2024, namun tidak menunjukkan pola yang konsisten. Sementara itu, nilai *Calinski-Harabasz Index* (CHI) pada beberapa tahun meningkat ketika K diperbesar, namun peningkatan tersebut tidak selalu sejalan dengan nilai *Silhouette*. Oleh karena itu, dengan mempertimbangkan konsistensi nilai *Silhouette* pada seluruh tahun pengamatan, $K=3$ dipilih sebagai jumlah kluster yang paling optimal untuk pemetaan kluster pembangunan kabupaten/kota di Provinsi Bali periode 2020–2024.

4.3. Evaluasi Kualitas Kluster

Evaluasi kualitas kluster dilakukan untuk menilai performa hasil pengelompokan kabupaten/kota di Provinsi Bali tahun 2020–2024 menggunakan nilai $K = 3$. Metrik yang digunakan dalam evaluasi adalah *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CHI). Hasil evaluasi kualitas kluster dapat dilihat pada Tabel 2.

Tabel 2. Evaluasi Kualitas Klaster

Tahun	Silhouette	DBI	CHI
2020	0.467339	0.597372	10.453485
2021	0.344673	0.834731	7.524924
2022	0.198522	0.725376	8.074470
2023	0.390162	0.723107	9.336803
2024	0.258808	0.657644	10.314022

Berdasarkan Tabel 2, kualitas klaster terbaik ditunjukkan pada tahun 2020 dengan nilai *Silhouette* 0,467, DBI 0,597, dan CHI 10,453, yang mengindikasikan klaster cukup kompak dan terpisah dengan baik. Pada tahun 2021 terjadi penurunan kualitas klaster ditandai dengan menurunnya *Silhouette* menjadi 0,345 dan meningkatnya DBI menjadi 0,835. Tahun 2022 menunjukkan kualitas klaster terendah dengan nilai *Silhouette* 0,199, meskipun nilai CHI masih berada pada 8,074. Selanjutnya, kualitas klaster kembali membaik pada tahun 2023, sementara tahun 2024 berada pada kategori sedang. Secara keseluruhan, metode VAE dan Improved K-Means mampu menghasilkan klaster yang cukup baik, dengan performa terbaik pada tahun 2020 dan 2023, serta performa terendah pada tahun 2022.

4.4. Hasil Clustering dan Interpretasi Klaster Pembangunan

Hasil clustering kabupaten/kota di Provinsi Bali periode 2020–2024 menggunakan metode VAE dan *Improved K-Means* menghasilkan pembagian wilayah ke dalam tiga klaster ($K=3$). Untuk memahami karakteristik masing-masing klaster, dilakukan analisis profil klaster menggunakan nilai rata-rata indikator IPM, PDRB, GR, TPT, dan P0 pada setiap tahun. Profil rata-rata indikator per tahun dan klaster dapat dilihat pada Tabel 3.

Tabel 3. Profil rata-rata klaster tiap tahun

Tahun	Cluster	IPM	PDRB	GR	TPT	P0
2020	0	72.99	43069	0.33	4.82	4.72
	1	82.76	78805	0.32	7.27	2.08
	2	70.95	34138	0.28	3.52	4.75
2021	0	74.24	44182	0.31	5.13	5.35
	1	82.93	75049	0.35	6.97	2.79
	2	68.36	29478	0.30	2.06	5.93
2022	0	71.62	39564	0.31	2.99	5.96
	1	79.83	59451	0.32	5.23	4.28
	2	82.13	99075	0.32	6.87	2.53
2023	0	74.71	49333	0.33	2.86	5.30
	1	83.90	100862	0.31	2.78	2.49
	2	71.95	39257	0.31	1.02	5.44
2024	0	72.84	45532	0.29	1.43	5.35
	1	83.87	132087	0.28	1.83	2.23
	2	81.15	69413	0.31	1.98	3.66

Nilai pada tabel tersebut menunjukkan bahwa komposisi indikator antar klaster memiliki variasi berbeda-beda pada setiap tahun, yang mencerminkan adanya perbedaan karakteristik pembangunan relatif antar wilayah. Selain itu, terlihat bahwa penomoran klaster (Cluster 0 – 2) tidak bersifat tetap dari tahun ke tahun dimana cluster 1 tidak selalu memiliki tingkat pembangunan IPM dan PDRB yang tinggi, sehingga interpretasi hasil pengelompokan perlu dilakukan dengan mengacu pada profil indikator.

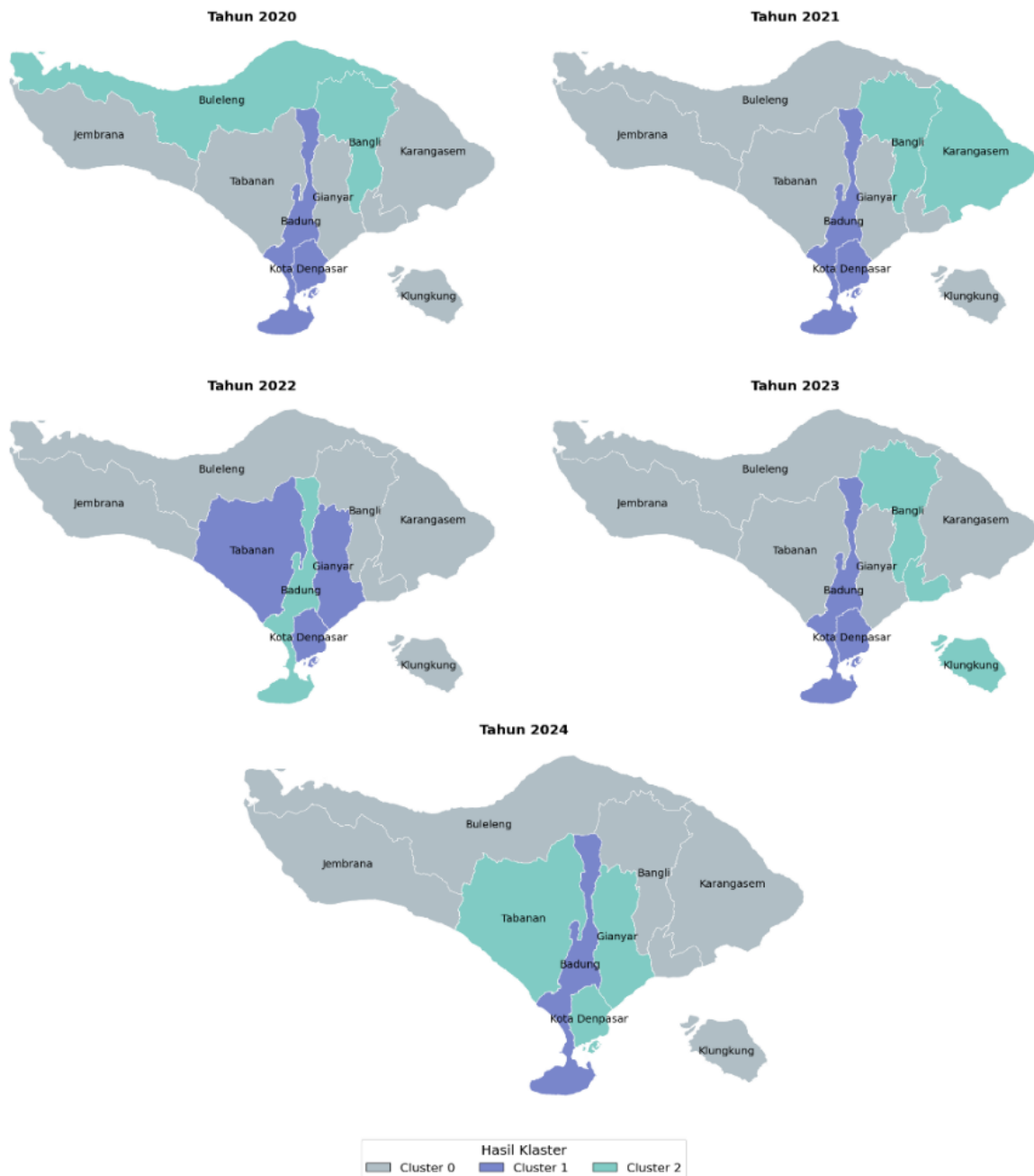
4.5. Interpretasi dan Pemetaan Klaster

Interpretasi hasil klaster dilakukan berdasarkan rata-rata profil tiap klaster yang disajikan pada Tabel 3, yang digunakan untuk membedakan tingkat pembangunan relatif antar klaster pada setiap tahun pengamatan. Klaster dengan nilai rata-rata indikator yang lebih tinggi merepresentasikan tingkat pembangunan yang lebih baik, sedangkan klaster dengan nilai rata-rata yang lebih rendah menunjukkan tingkat pembangunan yang relatif tertinggal. Perbedaan ini mencerminkan variasi kombinasi indikator IPM, PDRB per kapita, Gini Ratio (GR), Tingkat Pengangguran Terbuka (TPT), dan persentase penduduk miskin (P0) dalam membentuk karakteristik masing-masing klaster.

Berdasarkan profil tersebut, Klaster 1 secara umum menunjukkan tingkat pembangunan tertinggi, ditandai oleh IPM 79,83–83,90, PDRB per kapita 59.451–132.087, serta tingkat kemiskinan (P0) terendah 2,08–2,79. Klaster 0 merepresentasikan tingkat pembangunan menengah dengan IPM 71,62–74,71, PDRB per kapita 39.564–49.333, dan P0 4,72–5,96, sedangkan Klaster 2 menunjukkan tingkat pembangunan relatif lebih rendah dengan IPM yang fluktuatif 68,36–82,13, PDRB per kapita 29.478–69.413, dan P0 tertinggi 3,66–5,93. Namun, pada 2022 terjadi anomali, di mana Klaster 2 sementara merepresentasikan tingkat pembangunan tertinggi dengan IPM 82,13 dan PDRB per kapita 99.075, mengindikasikan adanya *cluster switching* yang bersifat temporer sebelum kembali menunjukkan pola dominan Klaster 1 pada periode berikutnya.

Pola perbedaan karakteristik klaster tersebut selanjutnya divisualisasikan dalam peta klaster pembangunan kabupaten/kota di Provinsi Bali periode 2020–2024, sebagaimana ditampilkan pada Gambar 3.

Peta Hasil Klaster Pembangunan Bali



Gambar 3. Peta persebaran anggota custer tiap tahun

Peta klaster menggambarkan persebaran spasial keanggotaan klaster pada setiap tahun pengamatan, di mana perbedaan warna menunjukkan keanggotaan klaster dan perubahan warna antar tahun mencerminkan dinamika perubahan klaster akibat variasi nilai indikator pembangunan. Karena proses klasterisasi dilakukan secara terpisah pada setiap tahun, penomoran klaster (Cluster 0–2) tidak bersifat tetap, sehingga interpretasi peta dilakukan dengan mengaitkan keanggotaan klaster terhadap profil indikator, bukan berdasarkan nomor klaster semata.

Hasil klasterisasi 2020–2021 menunjukkan pola yang relatif konsisten, dengan Badung dan Kota

Denpasar berada pada klaster dengan tingkat pembangunan tertinggi, ditandai oleh IPM tinggi 82,76–82,93, PDRB per kapita besar 75.049–78.805, serta tingkat kemiskinan (P0) terendah sebesar 2,08–2,79. Sebagian besar kabupaten lain di Bali tergolong dalam klaster pembangunan menengah dengan IPM 72,99–74,24 dan PDRB per kapita 43.069–44.182, sementara Bangli dan beberapa wilayah lainnya menunjukkan karakteristik pembangunan lebih rendah, dengan IPM 68,36–70,95, PDRB per kapita 29.478–34.138, serta tingkat kemiskinan P0 relatif tinggi 5,93.

Perubahan signifikan terjadi pada 2022, ketika Badung berpindah ke klaster dengan profil pembangunan tertinggi dengan IPM 82,13 dan PDRB per kapita sebesar 99.075, melampaui klaster yang sebelumnya, yang mengindikasikan adanya *cluster switching temporer*. Pada 2023–2024, struktur klaster kembali menunjukkan pola yang lebih stabil, dengan Badung sebagai wilayah paling maju, sementara Kota Denpasar, Gianyar, dan Tabanan berada pada klaster dengan IPM tinggi namun PDRB menengah, dan wilayah lainnya tetap menunjukkan karakteristik pembangunan menengah hingga rendah. Secara keseluruhan, temuan ini menegaskan bahwa pola klaster pembangunan antarwilayah bersifat dinamis, dipengaruhi oleh perubahan kinerja indikator sosial-ekonomi dari waktu ke waktu.

5. KESIMPULAN DAN SARAN

Penelitian ini mengelompokkan kabupaten/kota di Provinsi Bali periode 2020–2024 menggunakan *Variational Autoencoder* (VAE) dan *Improved K-Means* berdasarkan indikator IPM, PDRB per kapita, Gini Ratio (GR), Tingkat Pengangguran Terbuka (TPT), dan persentase penduduk miskin (P0). Hasil evaluasi menunjukkan bahwa $K = 3$ merupakan jumlah klaster yang paling optimal dalam merepresentasikan perbedaan karakteristik pembangunan antarwilayah. Berdasarkan profil rata-rata indikator, pada kebanyakan tahun Klaster 1 memiliki karakteristik pembangunan relatif lebih tinggi yang ditunjukkan oleh IPM dan PDRB per kapita yang lebih tinggi serta tingkat kemiskinan yang lebih rendah, sedangkan Klaster 0 dan Klaster 2 merepresentasikan karakteristik pembangunan menengah hingga lebih rendah. Hasil klasterisasi juga menunjukkan adanya pergeseran keanggotaan klaster antar tahun, yang mencerminkan dinamika perubahan kinerja sosial-ekonomi wilayah. Pengembangan penelitian selanjutnya dapat dilakukan dengan menambahkan indikator, membandingkan metode clustering, serta memperluas cakupan wilayah dan periode analisis.

DAFTAR PUSTAKA

- [1] H. Purmawan, I. Kustiani, and S. Sarkowi, "Keterpaduan Infrastruktur Terpadu Antar Sektor Untuk Mendukung Pengembangan Wilayah Kabupaten Pringsewu," *Jurnal Profesi Insinyur Universitas Lampung*, vol. 5, no. 1, pp. 7–10, Jun. 2024, doi: 10.23960/jpi.v5n1.118.
- [2] Fauzi Nur Kholis and T. Gunarto, "Analisis Pengaruh Kesenjangan Sosial, Indeks Pembangunan Manusia (IPM), dan Pengangguran Terhadap Kemiskinan di Provinsi Lampung (2012-2023)," *eCo-Buss*, vol. 7, no. 3, pp. 2128–2142, Apr. 2025, doi: 10.32877/eb.v7i3.2243.
- [3] S. Suraya, M. Sholeh, and U. Lestari, "Evaluation of Data Clustering Accuracy using K-Means Algorithm," *International Journal of Multidisciplinary Approach Research and Science*, vol. 2, no. 01, pp. 385–396, Dec. 2023, doi: 10.59653/ijmars.v2i01.504.
- [4] J. Pendidikan and D. Konseling, "Indeks Pembangunan Manusia Di Indonesia (Kajian Perubahan Metodologi Penghitungan) Universitas Pahlawan Tuanku Tambusai," vol. 5.
- [5] E. Erlin, E. Erni, M. Maggie, V. Vincent, and Z. Zidane, "Produk Domestik Regional Bruto Provinsi Sumatra Barat Menurut Lapangan Usaha," *Jurnal Penelitian Inovatif*, vol. 1, no. 2, pp. 187–192, Jan. 2022, doi: 10.54082/jupin.28.
- [6] Z. Zakiah, H. Isnaini, and M. Zuhrinal, "Kemiskinan dan Distribusi Pendapatan pada Masyarakat Pesisir di Kabupaten Aceh Tamiang," *JEEP*, vol. 4, no. 1, pp. 7–13, 2024, doi: 10.52300/jepp.v4i1.11755.
- [7] R. Wei and A. Mahmood, "Recent Advances in Variational Autoencoders with Representation Learning for Biomedical Informatics: A Survey," *IEEE Access*, vol. 9, pp. 4939–4956, 2021, doi: 10.1109/ACCESS.2020.3048309.
- [8] A. Singh and T. Ogunfunmi, "An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications," Jan. 01, 2022, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/e24010055.
- [9] A. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "Ikotun, Abiodun & Ezugwu, Absalom & AbualiK-means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data," *Inf. Sci. (N Y)*, vol. 662, 2022.
- [10] M. Idhom, A. Muhaimin, A. T. Damaliana, Trimono, M. Nasrudin, and S. S. M. Wara, "Evaluating Clustering Metrics for Time-Series Data: A Power Test Approach," in *2025 IEEE 11th Information Technology International Seminar (ITIS)*, 2025, pp. 368–373.
- [11] Hasan Yasir, "Pengukuran Silhouette Score dan Davies-Bouldin Index Pada Hasil Cluster K-Means dan DBSCAN," *Jurnal Informatika dan Teknik Elektro Terapan*, 2024.
- [12] A. N. Fitriani, K. Maulida Hindrayani, and T. Trimono, "ENHANCED CLUSTERING USING PSO-KMEDOIDS FOR GOVERNMENT AID DISTRIBUTION," *Jurnal Sistem Informasi dan Informatika (Simika) P-ISSN*, vol. 8, no. 2, pp. 2622–6901, 2025, [Online]. Available: <https://kalipuro.desa.id>
- [13] A. Idrus, N. Tarihoran, U. Supriatna, A. Tohir, S. Suwarni, and R. Rahim, "Distance Analysis Measuring for Clustering using K-Means and Davies Bouldin Index Algorithm," *TEM Journal*, vol. 11, no. 4, pp. 1871–1876, Nov. 2022, doi: 10.18421/TEM114-55.
- [14] E. Luthfi, A. Wahyu Wijayanto, and P. Statistika, "Analisis perbandingan metode

- hierarchical, k-means, dan k-medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia,” no. 4, pp. 761–773, 2021, [Online]. Available: <http://journal.feb.unmul.ac.id/index.php/INOVA-SI>
- [15] W. Nugraha, R. Sabaruddin, and S. Murni, “Teknik Scaling Menggunakan Robust Scaler Untuk Mengatasi Outlier Data Pada Model Prediksi Serangan Jantung Scaling Technique Using Robust Scaler to Overcome Data Outliers in the Heart Attack Prediction Model.” [Online]. Available: <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack>